

Roy Luo

437-363-4171 | royluocanada@gmail.com | <https://www.linkedin.com/in/luo-roy/> | <https://github.com/Depilicious>

EDUCATION

University of Waterloo

Statistics, Co-op

Sep 2023 – Apr 2027

Waterloo, Ontario

McMaster University

Mathematics and Statistics

Sep 2022 – Apr 2023

Hamilton, Ontario

TECHNICAL SKILLS

Languages: Python, R, SQL, C, Bash

Tools: Docker, Airflow, Terraform, dbt, GCP, PowerBI, Git, Spark, Snowflake

Libraries: pandas, NumPy, scipy, sklearn, Matplotlib, tidyverse, ggplot2

EXPERIENCE

Data Engineering Intern

Sep 2024 – Dec 2024

Questrade Financial Group

Toronto, ON

- Utilizing Terraform to secure and reduce developer friction of GCP services (Dataflow, Pub/Sub, BigQuery).
- Constructing a Python ETL pipeline with SQL that processed and migrated financial data from on-premises Oracle and SQL data sources to a BigQuery data warehouse.
- Enhancing the developer experience by restructuring the pretraining/datatools codebase, employing Docker for effective dependency management, and providing comprehensive documentation for each step in the process.

Data Engineering Intern

Jan 2024 – Apr 2024

Healthcare of Ontario Pension Plan

Toronto, ON

- Developed a set of management facing Power BI dashboards that provided insights into HR headcount data, using SQL and Apache Airflow to automate the underlying ETL pipelines and enable daily data refreshes.
- Wrote Python scripts to extract data from Workday REST APIs, increasing the amount of data available and providing greater insights into employee attrition rates.
- Developed SQL stored procedures to migrate data incrementally from an on-premises database to a Snowflake database, reducing reliance on legacy systems.

Data Analyst Intern

May 2023 – Aug 2023

Precision Insight Consulting

Toronto, ON

- Imported equity/bond trades using built-in Simcorp Dimension tools, mapping them to transaction codes and performing instrument type validation with a 99.5% accuracy rate.
- Configured and managed batch jobs for data extraction, optimizing data handling and processing efficiency and saving 15 hours per month of manual work.
- Constructed reference files documenting daily CAD/USD settlements in Excel and NAV reconciliation reports in Python, reducing reporting errors by 100% and enhancing data accessibility and accuracy.

PROJECTS

League of Legends Churn Analysis | *Python, pandas, sklearn*

- Leveraged a REST API to collect extensive gameplay data, including match statistics and player attributes for over 1,000 new players.
- Implemented and optimized machine learning models, including logistic regression, decision tree, random forests, and neural networks, achieving up to 78% accuracy in predicting player retention.
- Engineered relevant features and applied dimensionality reduction techniques like principal component analysis (PCA) to extract important patterns and reduce data complexity, achieving up to 95.75% variance explained.

Live Bitcoin Orderbook Dashboard | *R, SQL Server, PowerBI*

- Developed an R script to extract, clean and reformat live orderbook data from a websocket feed transmitting 25000+ Bitcoin orders a minute.
- Uploaded and stored data in SQL Server, which was loaded into PowerBI using DirectQuery, resulting in a 50% reduction in dashboard loading time compared to traditional data loading methods.
- Maintained a near real-time cryptocurrency orderbook dashboard with a refresh rate of 1 second, displaying important metrics like mid-market rate, order prices, market depth, and bid-ask spread.